ELSEVIER

Journal of Computational and Applied Mathematics 105 (1999) 129–139

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Algebraic approximants and the numerical solution of parabolic equations

George A. Baker, Jr.

Theoretical Division, Los Alamos National Laboratory, University of California, Los Alamos, NM 87545, USA

Received 26 June 1997; received in revised form 4 March 1998

Dedicated to Professor Haakon Waadeland on the occasion of his 70th birthday

Abstract

The heat equation is but one example of problems which involve multiple scales. There is a lot of transient behavior which for many problems is of no particular interest. What is of concern is the long time-scale behavior. However the presence of the short time-scale behavior would seem to require numerical integration methods to take very short time steps to follow the behavior accurately. For these problems, what is desired is a numerical method which is accurate for the long time-scale behavior, and causes the transients to die out quickly. That their rate of decay is not quite right is not important for this class of problems. A formalism is developed which allows the straightforward derivation of finite-difference schemes which involve several prior times from algebraic approximants. The algebraic approximants turn out to be, in a quite natural way, approximations to the function $\exp\{-4\mu[\arcsin(\sqrt{w/4})]^2\}$ where $\mu = \kappa\Delta t/(\Delta x)^2$ is the Courant number. Several of the simpler cases are investigated, and, of the implicit schemes, a couple are found to be, not only of higher order accuracy than most currently popular schemes, but also unconditionally stable and in fact unconditionally stiff stable! The higher order accuracy and stiff stability properties are just what is required for this sort of problem.

Keywords: Hermite–Padé approximant; Algebraic approximant; Padé approximant; Partial differential equation; Heat equation; Parabolic equation

1. Introduction and summary

An important problem in numerical analysis is the solution of the heat equation. This problem is, of course, only the simplest manifestation of a whole class of problems, but for the sake of clarity, I will confine my discussion to it. The basic problem is that there are a wide range of scales in the problem. For long times we know that all the transients decay to insignificance, but the long time scale behavior persists. Often it is just this long time behavior that is of interest and numerically we

0377-0427/99/$ - see front matter
PII: S0377-0427(99)00042-4

would like to take long time steps in order to get into this region quickly. This need suggests that we would like to have a high order method in order to preserve accuracy. In addition, at a minimum we want the method to be stable in the sense of Lax and Richtmyer. That is to say, as we know the true solution to the heat equation converges in time (given finite boundary conditions of course) we do not want the numerical solution to blow up. Even further, it is desirable that the short time scale transients decay as we know they should. This property is called stiff-stability or L-stable. We take as the heat equation, which is a parabolic equation taken here in one dimension for simplicity,

$$\frac{1}{\kappa}\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2}. \tag{1.1}$$

The traditional method of dealing with this problem is the method of lines where we replace $u(x_i,t) \to U_i(t)$ and obtain the 'space-discretized' equation,

$$\frac{1}{\kappa}\frac{\partial U_i}{\partial t} = \frac{U_{i+1} - 2U_i + U_{i-1}}{(\Delta x)^2} = \sum_{j=1}^{N} A_{i,j}U_j, \tag{1.2}$$

which defines the matrix $\mathscr{A}$. The solution to this equation is, by standard methods just

$$\boldsymbol{U}(t) = \exp(\mathscr{A}t)\boldsymbol{U}(0). \tag{1.3}$$

This solution of course suffers from the fact that the difference scheme used for the spatial derivative is only a limited order of accuracy approximation to the relevant derivatives, although it is accurate to all orders in t. Eq. (1.3) is explicitly for zero boundary conditions. The generalization is straightforward.

The traditional approach is to stick with the spatial difference scheme and to approximate the exponential function (of the matrix operator) for the case of a single time step, i.e., $t = \Delta t$. First we remark, as is well known, the simplest approximation is just $\exp(\mathscr{X}) = 1 + \mathscr{X}$. This leads to the usual explicit method. It is only stable for time steps limited by $0 < \kappa \Delta t a_j/(\Delta x)^2 \leqslant 2$, for all the eigenvalues a_j of the matrix $\mathscr{A}$. A more through investigation involves considering $\exp(\mathscr{A}\Delta t)$ as a function of Δt with matrix coefficients. The idea is to form 'operator Padé approximants' [2]. Since all the coefficients are functions of a single matrix, they all commute with each other and matters are considerable simpler than the general case. The Padé approximants $[M/M]$ [6], $[M-1/M]$ and $[M-2/M]$ [4] are all stable, and hence by power counting the $[M-1/M]$ and $[M-2/M]$ are also stiff stable. For this system, even though the same three-point, for example, spatial difference scheme is retained, when the denominator is a polynomial of degree M, the actual spatial difference scheme is a $(2M+1)$-point one with no increase in accuracy in the spatial derivative. It would be desirable, if we have to go to a higher-point difference scheme, to get better spatial accuracy out of it. It is a classic problem that the desirable conditions of (i) a small, number-of-points difference scheme, (ii) higher order accuracy, and (iii) stability, not to mention, a fortiori, stiff stability fight against each other.

Iserles [5] follows an alternative approach to the traditional one. In his approach the Courant number, $\mu = \kappa \Delta t/(\Delta x)^2$, is held constant and the approximation to the solution is expanded in terms of Δx. He concentrated his efforts on difference schemes which involve, as in the traditional methods just described, only one backward time. That is to say, they involve the times $u(x,t)$ and $u(x', t - \Delta t)$. There is interest in considering schemes which involve more than one prior time. In the second section, we develop a formalism for several prior times. This formalism is an extension

and modification of that of Iserles [5]. We show that the ideas involved correspond rather naturally to forming algebraic approximants [2] to the function $\exp\{-4\mu[\arcsin(\sqrt{w/4})]^2\}$.

In the third section we discuss several simple examples. The first example is a three-point, fifth-order accurate method. It is stiff stable. The second scheme is a five-point, seventh-order accurate method which is also stiff-stable. The last example is a three-point, seventh-order accurate method, which however is not uniformly stable in the Courant number, μ.

2. Methodology

Following the method of Iserles [5], we generalize to the case of $k+1$ backward time steps. The discretized version of Eq. (1.1) can be written as

$$\sum_{j=-\tau}^{\tau} \beta_j u_{l-j}^{(n)} = \sum_{m=0}^{k} \sum_{j_m=-\sigma_m}^{\sigma_m} \alpha_{m,j_m} u_{l-j_m}^{(n-m-1)}, \tag{2.1}$$

for a mesh $\Delta x = h$ and time step Δt. Let $\hat{u}(\theta)$ be the Fourier transform,

$$\hat{u}(\theta) = \sum_{l=-\infty}^{\infty} u_l \mathrm{e}^{\mathrm{i}l\theta h}, \quad -\frac{\pi}{h} < \theta < \frac{\pi}{h}, \tag{2.2}$$

$$u_l = \frac{h}{2\pi} \int_{-\pi/h}^{\pi/h} \hat{u}(\theta) \mathrm{e}^{-\mathrm{i}l\theta h} \, \mathrm{d}\theta. \tag{2.3}$$

If we Fourier transform the difference equation, we get,

$$\sum_{l=-\infty}^{\infty} \sum_{j=-\tau}^{\tau} \beta_j \mathrm{e}^{\mathrm{i}j\theta h} u_{l-j}^{(n)} \mathrm{e}^{\mathrm{i}(l-j)\theta h} = \sum_{m=0}^{k} \sum_{l=-\infty}^{\infty} \sum_{j_m=-\sigma_m}^{\sigma_m} \alpha_{m,j} \mathrm{e}^{\mathrm{i}j_m\theta h} u_{l-j_m}^{(n-m-1)} \mathrm{e}^{\mathrm{i}(l-j_m)\theta h}, \tag{2.4}$$

or

$$\sum_{m=-1}^{k} \tilde{Q}_{m,\sigma_m}(z) \hat{u}^{(n-m-1)}(\theta) = 0, \tag{2.5}$$

where $\sigma_{-1} = \tau$ and

$$\tilde{Q}_{-1,\sigma_{-1}} = \sum_{j=-\tau}^{\tau} \beta_j \mathrm{e}^{\mathrm{i}j\theta h}, \qquad \tilde{Q}_{m,\sigma_m}(z) = -\sum_{j=-\sigma_m}^{\sigma_m} \alpha_{m,j} \mathrm{e}^{\mathrm{i}j\theta h} \quad \text{for } 0 \leqslant m \leqslant k, \quad z = \mathrm{e}^{\mathrm{i}\theta h}. \tag{2.6}$$

If we Fourier transform the original differential equation (1.1), we get

$$\frac{\partial}{\partial t} \hat{u}(\theta, t) = -\kappa \theta^2 \hat{u}(\theta, t), \tag{2.7}$$

and therefore,

$$\hat{u}(\theta, t + \Delta t) = \mathrm{e}^{-\kappa\theta^2 \Delta t} \hat{u}(\theta, t). \tag{2.8}$$

We can rewrite, by the definition of the Courant number,

$$\mathrm{e}^{-\kappa\theta^2 \Delta t} = \mathrm{e}^{-\mu(\theta \Delta x)^2} = \mathrm{e}^{\mu(\ln z)^2} \tag{2.9}$$

Hence, if we want an approximation to Eq. (2.5) of the order of $(\Delta x)^p = h^p$, which is equivalent to $O(|z-1|^p)$, then we must have,

$$\left\{\sum_{m=-1}^{k} \tilde{Q}_{m,\sigma_m}(z)\left[e^{-\mu(\ln z)^2}\right]^{m+1}\right\}\hat{u}^{(n)}(\theta) = c|z-1|^{p+1} + O(|z-1|^{p+2}), \quad c \neq 0, \tag{2.10}$$

by the properties of the exponential function. The order of approximation to the differential equation (1.1) is $p-2$.

Let us define

$$f(z) = e^{-\mu(\ln z)^2} \tag{2.11}$$

The function $f(z)$ has the property that $f(z) = f(z^{-1})$ because $\ln z = -\ln z^{-1}$ and $f(z)$ depends only on the square of $\ln z$. Further, if we explicitly display the parametric dependence, then $f(z,\mu) = [f(z,-\mu)]^{-1}$ by the properties of the exponential function. By way of reference, Iserles [5] gives the initial terms of the expansion,

$$\begin{aligned} f(z) = 1 &- \mu\zeta^2 + \mu\zeta^3 + \left(-\frac{11}{12}\mu + \frac{1}{2}\mu^2\right)\zeta^4 + \left(\frac{5}{6}\mu - \mu^2\right)\zeta^5 + \left(-\frac{137}{180}\mu + \frac{17}{12}\mu^2 - \frac{1}{6}\mu^3\right)\zeta^6 \\ &+ \left(\frac{7}{10}\mu - \frac{7}{4}\mu^2 + \frac{1}{2}\mu^3\right)\zeta^7 + \left(-\frac{363}{560}\mu + \frac{967}{480}\mu^2 - \frac{23}{24}\mu^3 + \frac{1}{24}\mu^4\right)\zeta^8 \\ &+ \left(\frac{761}{1260}\mu - \frac{89}{40}\mu^2 + \frac{3}{2}\mu^3 - \frac{1}{6}\mu^4\right)\zeta^9 + \cdots, \end{aligned} \tag{2.12}$$

where $\zeta = z - 1$. It is clear from the definition that, the coefficients of $\zeta^j \; \forall j \geqslant 2$ are all proportional to μ. Further, Iserles has proven that they are polynomials of exact degree $[j/2]$.

The $z \to 1/z$ invariance property ensures that when we solve Eq. (2.10) for the α's and the β's that in Eq. (2.1) $\beta_j = \beta_{-j}$ for $0 \leqslant j \leqslant \tau$ and $\alpha_{m,j} = \alpha_{m,-j}$ for all $0 \leqslant j \leqslant \sigma_m$ with $0 \leqslant m \leqslant k$. This result corresponds to the $x \to -x$ symmetry of the original parabolic differential equation (1.1). Because of this coefficient equality, we have only the combinations $z^j + z^{-j}$ appearing. This quantity is just $2\cos(j\theta h)$ by the standard identities. As it is well known that $\cos(j\theta h)$ is a polynomial in $\cos(\theta h) = z + 1/z$, we may clearly re-express the Q's as polynomials in the quantity $z + 1/z$. It is convenient next to shift the origin by 2 and so we can write the Q's as polynomials in $w = 2 - z - 1/z = -\zeta^2/(1+\zeta)$. The series expansion (2.12) can be rewritten in powers of w. The existence of this transformation $w(\zeta)$ explains, via Theorem 3 of Baker et al. [1] on the structure of the Padé table for functions of variables of this form, the 2×2 blocks along the diagonal of the Padé table for $f(\zeta)$ found by Iserles [5]. To see that the series (2.12) can be rewritten in powers of w, notice that $w = -(z^{1/2} - z^{-1/2})^2 = 4\sin^2(\theta h/2)$. Since $(\ln z)^2 = -\theta^2 h^2$ we can write $(\ln z)^2 = -4[\arcsin(\sqrt{w/4})]^2$. Now it is well known that $\arcsin x$ is an odd series in x. Thus we conclude that $(\ln z)^2$ is an even series in $\sqrt{w}$ or in other words, it is a series in integer powers of w. Hence the conclusion is that $f(\zeta)$ can be written as a power series in w. The initial terms are

$$f(w) = 1 + \mu w + \left(\frac{1}{12}\mu + \frac{1}{2}\mu^2\right)w^2 + \left(\frac{1}{90}\mu + \frac{1}{12}\mu^2 + \frac{1}{6}\mu^3\right)w^3 + \cdots. \tag{2.13}$$

Note is taken that as $f(w)=\exp[\mu\{w+\mathrm{O}(w^2)\}]$ the highest power of μ in the coefficient of w^n is $\mu^n/n!$. We may now recast Eq. (2.10) as

$$\sum_{m=-1}^{k} Q_{m,\,\sigma_m}(w)[f(w)]^{m+1}=c|z-1|^{p+1}+\mathrm{O}(|z-1|^{p+2}),\quad c\neq 0, \tag{2.14}$$

where $Q_{m,\,\sigma_m}(w)$ is a polynomial in w of degree σ_m. If we select, as $w=\zeta^2+\cdots$,

$$2k+1+2\sum_{m=-1}^{k}\sigma_m=p \tag{2.15}$$

then we recognize (2.14)–(2.15) as the defining equations for the polynomials which define the algebraic approximant to $f(w)$ of degree k when the expansion is taken about $w=0$ [2]. The algebraic approximants themselves are defined by the solution of the equation

$$\sum_{m=-1}^{k} Q_{m,\,\sigma_m}(w)y(w)^{m+1}=0, \tag{2.16}$$

subject to the initial conditions,

$$y^{(j)}(0)=f^{(j)}(0),\quad j=0,\ldots,k-1, \tag{2.17}$$

in the usual case and are denoted by

$$\langle\sigma_{-1}/\sigma_0,\ldots,\sigma_k\rangle=\langle\sigma_{-1}/\boldsymbol{\sigma}\rangle=y(z). \tag{2.18}$$

The defining equation (2.14) for the algebraic polynomials continues to hold if we reverse the sign of μ. By use of the property $f(w,-\mu)=[f(w,\mu)]^{-1}$ and multiplication by $[f(w,-\mu)]^{k+1}$ we find by comparison with Eq. (2.14), in the case where $\sigma_j=\sigma_{k-j-1}$, the property

$$Q_{j,\,\sigma_j}(w,\mu)=Q_{k-j-1,\,\sigma_{k-j-1}}(w,-\mu),\quad -1\leqslant j\leqslant k. \tag{2.19}$$

Note that the case of all equal σ's is included here. It follows directly from this result that $y(w,\mu)=[y(w,-\mu)]^{-1}$.

Next we report on the Q's as a function of μ. For notational convenience we write $[f(w)]^{1+j}=\sum_{m=0}^{\infty} f_m^{(j+1)}w^m$. Then by Cramer's rule, the solution of Eq. (2.14) is

$$Q_{j,\,m_j}(z)=\det\begin{vmatrix}
f_0^{(-1)} & 0 & \ldots & 0 & & f_0^{(j)} & 0 & \ldots & 0 & \\
f_1^{(-1)} & f_0^{(-1)} & \ldots & 0 & & f_1^{(j)} & f_0^{(j)} & \ldots & 0 & \\
\vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots & \\
f_{\sigma_{-1}}^{(-1)} & f_{\sigma_{-1}-1}^{(-1)} & \ldots & f_0^{(-1)} & \ldots & f_{\sigma_j}^{(j)} & f_{\sigma_j-1}^{(j)} & \ldots & f_0^{(j)} & \ldots \\
\vdots & \vdots & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots & \\
f_s^{(-1)} & f_{s-1}^{(-1)} & \ldots & f_{s-\sigma_{-1}}^{(-1)} & & f_s^{(j)} & f_{s-1}^{(j)} & \ldots & f_{s-\sigma_j}^{(j)} & \\
0 & 0 & \ldots & 0 & \ldots & 1 & z & \ldots & z^{\sigma_j} & \ldots
\end{vmatrix}, \tag{2.20}$$

where

$$s=\sum_{j=-1}^{k}(\sigma_j+1)-2. \tag{2.21}$$

In terms of this explicit solution, since every coefficient in Eq. (2.13) is a polynomial in μ, it must be that all the coefficients of the algebraic polynomials are also polynomials. Since all the coefficients in Eq. (2.13), except the coefficient of w^0 is divisible by μ, it must be that every algebraic polynomial coefficient, and hence the entire equation is divisible by $\mu^{\tilde{s}}$, where $\tilde{s} = s - \max_{-1\leqslant m\leqslant k}(\sigma_m)$. The largest power of μ that can occur is $s(s+1)/2$, and it only occurs when $s = k$.

It is useful at this point to introduce the definitions of stability and stiff-stability.

Definition 2.1. A difference scheme is vN-acceptable for a given value of μ if the multiplication factor for advancing by one-time step satisfies the von Neumann condition,

$$|y_j(z = \mathrm{e}^{\mathrm{i}\theta}, \mu)|^{-1} \leqslant 1, \quad 0 \leqslant \theta \leqslant 2\pi, \quad 0 \leqslant j \leqslant k, \tag{2.22}$$

for each of the $k+1$ solutions y_j of (2.16).

Note that by use of standard arguments it can be shown that vN-acceptability implies stability in the sense of Lax and Richtmyer. That is, it implies for well posed problems that there exists, independent of Δt but dependent on the initial conditions, an upper bound $M(t)$ on $|u(x,t)|$. Well posed includes in this case the condition that the initial conditions (on the mesh points) $u(x, n\Delta t)$ for $0 \leqslant n \leqslant k$ may be represented in the form

$$\hat{u}^{(n)}(\theta) = \sum_{j=0}^{k} a(\theta, j)[y_j(\theta)]^{-n}, \quad 0 \leqslant n \leqslant k, \tag{2.23}$$

where all the $a(\theta, j)$ are finite. Since for each θ the solution of these equations involves a Van der Monde determinant in the denominator, a sufficient condition for this representation to hold is that for each θ the $y_j(\theta)$ be distinct [and of course the $u(x, n\Delta t)$ be finite].

Definition 2.2. A difference scheme is L-stable for a given value of μ if it is vN-acceptable and in addition the high-frequency components decay in time. That is, $|y_j(z = -1, \mu)|^{-1} < 1$, for each of the $k+1$ solutions of Eq. (2.16).

From the Fourier transform solution we expect that

$$y(z = -1, \mu)^{-1} = \mathrm{e}^{-\pi^2\mu} \underset{\mu\to\infty}{\to} 0. \tag{2.24}$$

Iserles [5] has proven for the case $k = 0$, i.e., ordinary Padé approximants, that for schemes based on $\langle\sigma/\sigma\rangle(\equiv[\sigma/\sigma])$ the difference schemes are vN-acceptable for all σ. They are not, however L-stable.

3. Examples

To illustrate the above formalism, I have computed the algebraic polynomials for three of the simplest approximants to $f(w)$ (2.13). They are (i) for the $\langle 1/0, 0\rangle$, after normalization,

$$Q_{-1}(w) = 1 + \frac{12\mu^2 w}{1+18\mu}, \qquad Q_0(w) = -\left(\frac{2+24\mu}{1+18\mu}\right), \quad Q_1(w) = \frac{1+6\mu}{1+18\mu}. \tag{3.1}$$

(ii) For the $\langle 2/0,0\rangle$ after normalization,

$$\begin{aligned} Q_{-1}(w) &= 1+\left(\frac{\frac{1}{6}\mu+\mu^2}{\frac{1}{90}+\frac{1}{4}\mu+\frac{7}{6}\mu^2}\right)\mu w+\left(\frac{\frac{1}{360}+\frac{1}{12}\mu+\frac{1}{3}\mu^2}{\frac{1}{90}+\frac{1}{4}\mu+\frac{7}{6}\mu^2}\right)\mu^2 w^2,\\ Q_0(w) &= -\left(\frac{\frac{1}{45}+\frac{1}{3}\mu+\frac{4}{3}\mu^2}{\frac{1}{90}+\frac{1}{4}\mu+\frac{7}{6}\mu^2}\right), \quad Q_1(w)=\left(\frac{\frac{1}{90}+\frac{1}{12}\mu+\frac{1}{6}\mu^2}{\frac{1}{90}+\frac{1}{4}\mu+\frac{7}{6}\mu^2}\right). \end{aligned} \tag{3.2}$$

(iii) For the $\langle 1/1,0\rangle$ after normalization,

$$\begin{aligned} Q_{-1}(w) &= 1-\left(\frac{\frac{1}{60}-\mu^2}{\frac{1}{40}+\frac{1}{2}\mu+\frac{5}{2}\mu^2}\right)\mu w,\\ Q_0(w) &= -\left(\frac{\frac{1}{20}+\frac{1}{2}\mu+2\mu^2}{\frac{1}{40}+\frac{1}{2}\mu+\frac{5}{2}\mu^2}\right)+\left(\frac{\frac{1}{60}+\frac{1}{2}\mu+2\mu^2}{\frac{1}{40}+\frac{1}{2}\mu+\frac{5}{2}\mu^2}\right)\mu w,\\ Q_1(w) &= \frac{\frac{1}{40}-\frac{1}{2}\mu^2}{\frac{1}{40}+\frac{1}{2}\mu+\frac{5}{2}\mu^2}. \end{aligned} \tag{3.3}$$

According to Eq. (2.15), Eq. (3.1) is fifth order accurate, and both Eqs. (3.2) and (3.3) are seventh order accurate. It turns out that Eq. (3.3) is not vN-acceptable uniformly in μ, as can be seen by noticing that in the large μ limit, the solution can grow by about a factor of two at each time step. I will not consider it further. The case $\langle 1/0,0\rangle$ was considered by Baker [3]. I will summarize some of those results for this case. First,

Theorem 3.1. *The difference scheme defined by Eq.* (3.1) *is L-stable for* $\mu > 0$.

He reports that the usual truncation analysis of the defining equation, (2.14) for this scheme gives a leading order error of $[\frac{1}{40}\Delta t(\Delta x)^4+\frac{1}{18}(\Delta t)^2(\Delta x)^2+\frac{1}{3}(\Delta t)^2]\theta^6$. In addition he has compared it numerically with the Crank-Nicolson method and the Crandall method which are of order of accuracy 3 and 5 respectively. There are basically three modes of comparison to be used. One is the time scale. That is if we normalize $\mu=\tilde{\mu}$ so that the decay of the slowest mode equals that for the exact solution for a given value of μ, how closely does $\tilde{\mu}=\mu$? The second parameter is how accurately does the method reproduce the asymptotic value of the amplitude of the dominant (slowest to decay) mode? The third criterion is how accurately does the (time normalized) solution track the exact solution? As a sample case he use a bar of uniform initial unit temperature with boundary temperatures of zero. He used 49 interior spatial points. He found that the Crandall method to be somewhat better on the time scale over the range checked ($\mu=1,\ 100$) and the Crank-Nicolson to be worse for $\mu<10$ but better for $\mu>10$. He found that on the second and third criteria, that the quadratic approximant scheme was either indistinguishable (small μ) or uniformly superior over the rest of the range checked (up to $\mu=1000$). Note that both the Crank-Nicolson and the Crandall method fail by divergence on the scale of the exact solution for $\mu>15$ or 20, while the quadratic approximant method does not. For $\mu=1000$ the quadratic approximant method develops oscillations with time step, but still does not diverge on the scale of the exact solution.

The finite difference scheme (3.2) is seventh-order accurate in contrast the scheme (3.1) which is fifth order accurate. However it is a five-point difference scheme, by which I mean there are five spatially different points at the new time involved in each equation. The scheme (3.1) is a three point scheme in this sense. The point is that in order to solve these equations for the values at the new time, one must, for a three point scheme, invert a tridiagonal matrix. There are very quick algorithms available to compute this solution. The five-point scheme leads to a band diagonal matrix with 2 non-zero supra-diagonals and two non-zero sub-diagonals. There are also quick methods to deal with this case, but not quite so nice as with the pure tridiagonals. Also, the higher point schemes raise some boundary condition issues which we will not discuss. I have obtained the following results for the finite-difference scheme (3.2).

Theorem 3.2. *The difference scheme defined by Eq.* (3.2) *is L-stable for* $\mu > 0$.

Proof. The two solutions of Eq. (2.16) for the case $k = 1$ are by the quadratic formula just,

$$\mathscr{Y}_{\pm}(w,\mu) = [y_{\mp}(w,\mu)]^{-1} = \frac{-Q_0(w) \pm \sqrt{Q_0(w)^2 - 4Q_{-1}(w)Q_1(w)}}{2Q_{-1}(w)}. \tag{3.4}$$

For the case of Eq. (3.2) we get

$$\mathscr{Y}_{\pm} = \frac{\left(\begin{array}{l} 8 + 120\mu + 480\mu^2 \pm \\ \sqrt{(60\mu)^2(1+6\mu)^2 - 8(2+15\mu+30\mu^2)[(60\mu+360\mu^2)\mu w + (1+30\mu+120\mu^2)(\mu w)^2]} \end{array}\right)}{2[4 + 90\mu + 420\mu^2 + (60\mu + 360\mu^2)\mu w + (1 + 30\mu + 120\mu^2)(\mu w)^2]}. \tag{3.5}$$

For the case of $w = 0$, we can compute directly that

$$\mathscr{Y}_{+} = 1, \quad \frac{1}{7} \leqslant \mathscr{Y}_{-} = \frac{2 + 15\mu + 30\mu^2}{2 + 45\mu + 210\mu^2} = 1 - \frac{30\mu + 180\mu^2}{2 + 45\mu + 210\mu^2} \leqslant 1. \tag{3.6}$$

As long as the discriminant is non-negative, in $\mathscr{Y}_{+}$ the numerator decreases and the denominator increases as w increases, so in this region the condition $1 \geqslant \mathscr{Y}_{+} > \mathscr{Y}_{-} \geqslant 0$ remains valid. The last inequality follows from the theory of equations as both the sum and the product of the two roots are positive, and the positivity of $\mathscr{Y}_{+}$.

The maximum value of w is 4. The value of the discriminant for $w = 4$ is

$$-496\mu^2 - 18240\mu^3 - 192960\mu^4 - 691200\mu^5 - 460800\mu^6 \leqslant 0. \tag{3.7}$$

Thus it is always the case for $w = 4$ ($z = -1$) that the two factors $\mathscr{Y}_{\pm}$ are complex conjugates. In this case,

$$|\mathscr{Y}_{\pm}(w)|^2 = \frac{Q_1(w)}{Q_{-1}(w)}. \tag{3.8}$$

By inspection of Eqs. (3.2) and (3.8) we see that $|\mathscr{Y}_{\pm}(w)|^2$ is a decreasing function of w and that it is continuous at the point where the discriminant vanishes. Thus $|\mathscr{Y}_{\pm}(w)|^2 \leqslant 1$ by our previous results when there were two real square roots. Thus the scheme is vN-acceptable. It is of interest to

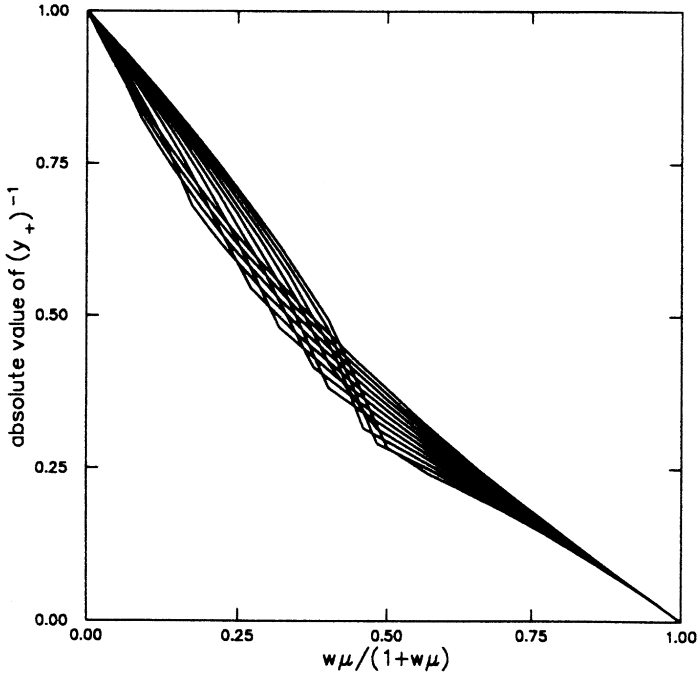

Fig. 1. The larger factor for scheme (3.2). The lines are for various values of w.

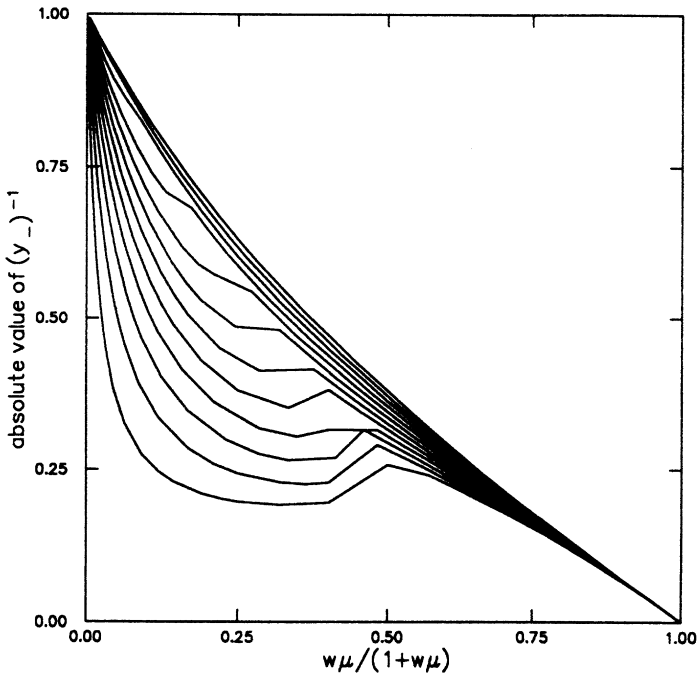

Fig. 2. The smaller factor for scheme (3.2). The lines are for various values of w.

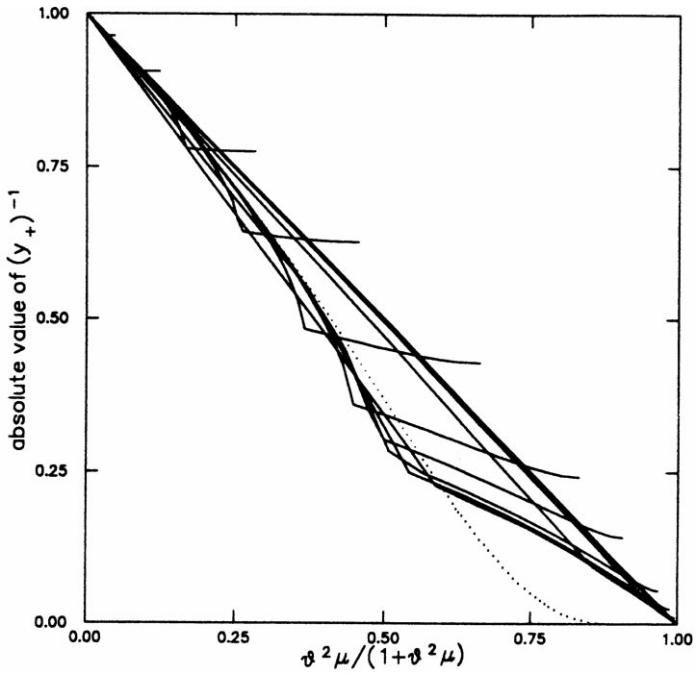

Fig. 3. The larger factor for scheme (3.2). Each solid line corresponds to a constant value of μ. The dotted line shows the exact result (2.11) for the factor.

note that for the case $w = 4$,

$$|\mathscr{y}_{\pm}(4)|^2 = \frac{4 + 30\mu + 60\mu^2}{4 + 330\mu + 1876\mu^2 + 480\mu^3 + 6720\mu^4}$$

$$= 1 - \frac{300\mu + 1816\mu^2 + 480\mu^2 + 6720\mu^4}{4 + 330\mu + 1876\mu^2 + 480\mu^3 + 6720\mu^4} < 1, \tag{3.9}$$

when $\mu > 0$. Hence the scheme is also L-stable. □

We note that, asymptotically as $\mu \to \infty$ $|\mathscr{y}_{\pm}(4)| \approx 0.1788/\mu$, which is better than the result [3] for the scheme (3.1), $|\mathscr{y}_{\pm}(4)| \approx 0.3536/\sqrt{\mu}$, as the exact result decays as $\exp(-\mu\pi^2)$.

In order to illustrate these results we show in Figs. 1 and 2 the values of the larger $(y_+)^{-1}$ and the smaller $(y_-)^{-1}$ factors, respectively. The obvious break in the slopes occurs where the two real roots become a complex conjugate pair. The lines are for constant values of w with μ changing.

It is also of interest is to plot curves for fixed μ as w changes. I have done so in Fig. 3. It is to be noticed that monotonic decrease property mentioned above holds. Again there is a noticeable break in the slope where the two real roots become a complex conjugate pair. The exact result is plotted as a dotted line for comparison. The range of values of μ is from 10^{-3} to 5×10^6. The results for the very largest values of μ are the straight diagonal lines. For smaller values of μ the results track the exact results closely, then deviate in the downward direction and finally change to a relatively flat curve where there are two complex conjugate pairs. The reason that the curves

stop short of the right hand edge is that $\theta \leqslant \pi$ and the right edge corresponds to an infinite value of $\theta^2 \mu$.

References

[1] G.A. Baker, Jr., J.L. Gammel, J.G. Wills, An investigation of the applicability of the Padé approximant method, J. Math. Anal. & Appl. 2 (1961) 405–418.
[2] G.A. Baker, Jr., P. Graves-Morris, in: G.-C. Rota (Ed.), Padé Approximants, 2nd ed., Encyclopedia of Mathematics and its Applications, vol. 59, Cambridge University Press, New York, 1996.
[3] G.A. Baker, Jr., Stable, implicit differencing schemes for the heat equation, Los Alamos National Lab, preprint.
[4] B.L. Ehle, *A*-stable methods and Padé approximants to the exponential, SIAM J. Math. Anal. 4 (1973) 671–680.
[5] A. Iserles, Order stars, approximations and finite differences III, SIAM J. Math. Anal. 16 (1985) 1020–1033.
[6] R.S. Varga, On higher order stable implicit methods for solving parabolic partial differential equations, J. Math. Phys. 40 (1961) 220–231.